DOCUMENT RESUME

ED 227 355 CE 035 458

AUTHOR	Armijo, Louis A.; Appleby, Judith A.
TITLE	Developing Vocational Competency Tests. Module 19. [Vocational Education Curriculum Specialist.]
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, DC.
PUB DATE	82
CONTRACT	300-79-0735
NOTE	54p.; For related documents, see ED 215 114-132 and CE 035 456-459.
AVAILABLE FROM	East Central Network, Sangamon State University, E-22, Springfield, IL 62708 ($4.00; complete set--20 modules, instructor's guide, audio cassette, field test report--$45.00).
PUB TYPE	Guides - Classroom Use - Materials (For Learner) (051)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Behavioral Objectives; Competence; Competency Based Education; Criterion Referenced Tests; *Disabilities; Inservice Education; *Job Skills; Learning Activities; Learning Modules; Performance Tests; Postsecondary Education; Secondary Education; Student Evaluation; *Test Construction; *Testing; *Vocational Education; Vocational Education Teachers
IDENTIFIERS	*Competency Tests; *Curriculum Specialists

ABSTRACT

This module, the third of four units about vocational competency measurement, is module 19 in the Vocational Education Curriculum Specialist series. The purpose stated for the document is to help in the development of both written and performance competency tests based on the needs identified for testing a particular program and the standards and priorities established for job-related tasks. Content is organized into four sections, each of which focuses on one goal and two or more objectives. Section 1 summarizes important considerations in vocational competency-test development, including test validity, reliability, practicality, and testing individuals with special needs. In section 2, important considerations in designing initial test specifications are overviewed. Sections 3 and 4 discuss the critical tasks in developing paper-and-pencil and performance tests. Each section concludes with individual study activities, discussion questions, and group activities. Self-check items and possible responses to them are appended for use as a pretest and review of the module content. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227355

DEVELOPING VOCATIONAL COMPETENCY TESTS

Module 19

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Louis A. Armijo

with the assistance of Judith A. Appleby

CE035458

Developed by the American Institutes for Research under support from the Office of Vocational and Adult Education, U.S. Department of Education. 1982.

The VECS materials are printed and distributed by
East Central Network for Curriculum Coordination
Sangamon State University, E-22
Springfield, IL 62708
217/786-6375

The information reported herein was obtained pursuant to Contract No. 300-79-0735 with the U.S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to document information according to their observation and professional judgment. Consequently, information, points of view, or opinions stated do not necessarily represent official Department of Education position or policy.

Table of Contents

	Page
Introduction	3
Overview	3
Instructions to the Learner	4
Goals and Objectives	5
Resources	6
GOAL 1	7
What Are Important Considerations In Test Development?	9
Test Validity	9
Test Reliability	9
Test Practicality	10
Testing Individuals With Special Needs	10
Individual Study Activities	13
Discussion Questions	13
Group Activity	13
GOAL 2	15
What Do You Include In Initial Test Specifications?	17
Types of Measures to be Used and Their Formats	17
Individual Study Activities	23
Discussion Questions	23
Group Activity	23

Page

GOAL 3 25

How Do You Develop Paper-and-Pencil Tests? 27

Create An Item Budget 27

Assign Test Items 29

Review and Modify Items 32

Balance the Key 33

Prepare Test Administration Instructions 34

Review and Revise (The Second Time) 35

Conduct a Pilot Test 36

Conduct a Field Test 36

Analyze Field Test Findings 37

Review the Test 37

Individual Study Activities 39

Discussion Questions 39

Group Activity 39

GOAL 4 41

How Do You Develop Performance Tests? 43

Select the Tasks to be Tested 43

Structure the Tasks 44

Prepare Administration Instructions 45

Develop Rating Sheets 45

Review by Consultants 45

Test the Performance Test 48

Page

Individual Study Activities 50

Discussion Questions 50

Group Activity 51

Summary 53

Appendices 57

Self-Check 59

Self-Check Responses 61

Recommended References 65

Tables

		Page
Table 1.	Suggested Adaptation Procedures for Special Groups	12
Table 2.	Components of a Multiple-Choice Item . . .	20
Table 3.	Suggested Instructions to Test Item Writers	31
Table 4.	Areas of Concern When Reviewing Paper-and-Pencil Test Items	33
Table 5.	Suggested Content Areas of an Examiner's Test Manual	35
Table 6.	Checkpoints Used in the Development of a Performance Test for Diesel Mechanic	46
Table 7.	A Sample Rating Sheet for Diesel Mechanic	47
Table 8.	Checklist for Reviewing Performance Test Items	48

ACKNOWLEDGMENTS

The discussions and techniques presented in Modules 17 through 20 of the VECS series are based on the work of the American Institutes for Research in carrying out the Vocational Competency Measures (VCM) project under contract with the Office of Vocational and Adult Education, U. S. Department of Education. The project was a major effort, beginning in October 1979 and continuing through 1982, to provide a national model for vocational competency test development.

The VCM project had four major objectives:

(1) To develop competency tests in selected occupations, representing each of the seven major areas: trade and industry, home economics, health, distributive education, technical, business and office, and agriculture;

(2) To establish their usefulness through extensive field testing and evaluation;

(3) To promote their acceptance and use in vocational education programs;

(4) To design and help implement a program for continuing occupational competency test development on a self-supporting basis.

The successful implementation of the project was due to the efforts of many people. Senior project staff responsible for specific tasks were:

Dr. Albert B. Chalupsky, Project Director
Ms. Marion F. Shaycoft, Director of Sampling and Test Quality Control, and Test Team Leader
Dr. Malcolm N. Danoff, Director of Field Coordination and Validation
Dr. Robert A. Weisgerber, Director of Competency Requirements Analysis and Test Team Leader
Ms. Judith A. Appleby, Director of Dissemination
Dr. John G. Claudy, Test Team Leader
Dr. William S. Farrell, Jr., Test Team Leader
Dr. John Caylor, Test Team Leader
Dr. Louis A. Armijo, Field Coordinator
Ms. Marie R. Peirano, Field Coordinator
Ms. Jeanette D. Wheeler, Test Editor and Production Coordinator

Mr. Steven Zwillinger, Department of Education Project Officer, provided support to staff throughout the project.

Introduction

This is the third of four modules dealing with the use, development, and validation of vocational competency tests. Earlier modules provided an overview of using competency measures in vocational education programs (Module 17) and a discussion of how to determine requirements for vocational competency measures (Module 18). The last module (Module 20) considers approaches to validating competency tests and using test results.

The purpose of this module is to help you develop competency tests--both written and performance--based on the needs you've identified for testing a particular program and the standards and priorities you've established for job-related tasks. The techniques presented here are based on the experiences of the American Institutes for Research in conducting the Vocational Competency Measures (VCM) project for the U.S. Department of Education as well as on previous test development experience of project staff.

Overview

The development of a vocational competency test requires the planning, coordination, and skillful execution of many activities. A test that has been developed following the procedures outlined in this module should provide supervisors, instructors, and students with information on how closely the skills taught and learned in the educational program compare with the work standards and skills expected in industry.

The approach used in this module provides considerable flexibility in the development process, but at the same time, it has a sufficiently structured framework to provide clear guidance. Although the test development procedures are intended for use in moderate to large test development efforts, small districts or individual schools will also find information that can be adapted to smaller-scale projects. If you want a more detailed knowledge of competency testing or testing in general, the Recommended References in the Appendices should be useful.

The development of a vocational competency test is both a creative and a mechanical process. This module can only describe the mechanics of a test development project. It is hoped that the framework given will allow you to use your creative talents most fully and effectively.

Instructions to the Learner

The Self-Check items and possible responses to them are found in the Appendices. These questions have two purposes. First, before you begin work on the module, you may use them to check quickly whether you have already learned the information in previous classes or readings. In some instances, with the consent of your instructor, you might decide to skip a whole module or parts of one. The second purpose of the Self-Check is to help you review the content of modules you have studied in order to assess whether you have achieved the module's goals and objectives.

You can also use the list of goals and objectives that follows to determine whether the module content is new to you and requires in-depth study, or whether the module can serve as a brief review before you continue to the next module.

Goals and Objectives

Goal 1: Summarize important considerations in vocational competency test development.

Objective 1.1 State the importance of test validity, test reliability, and test practicality.

Objective 1.2 List important considerations in developing tests to include individuals with special needs.

Goal 2: Summarize important considerations in designing initial test specifications.

Objective 2.1 State the purpose of designing initial test specifications and list items to be included in the specifications.

Objective 2.2 Compare the strengths and weaknesses of paper-and-pencil tests and performance tests.

Objective 2.3 Describe common formats of paper-and-pencil tests.

Objective 2.4 Describe types of performance evaluation.

Goal 3: Discuss the critical tasks in developing paper-and-pencil tests.

Objective 3.1 Describe the importance and process of creating an item budget.

Objective 3.2 List important considerations in the initial review and modification of test items.

Objective 3.3 Compare and contrast the processes of pilot testing and field testing.

Objective 3.4 Based on field testing, discuss the bases for revising test items.

Goal 4: Discuss the critical tasks in developing performance tests.

Objective 4.1 List the components of a performance test.

Objective 4.2 Identify key considerations in selecting and structuring tasks for performance test development.

Objective 4.3 Describe key considerations in reviewing performance test items.

Resources

In order to complete the learning activities in this module, you will need information contained in the following publication:

Erickson, R. C., & Wentling, T. L. Measuring student growth: Techniques and procedures for occupational education. Urbana, Ill.: Griffon Press, 1976.

GOAL 1: Summarize important considerations in vocational competency test development.

What Are Important Considerations in Test Development?

The development of any test requires that certain technical and practical considerations be kept in mind. The test developer should attempt to have the final test satisfy as closely as possible the three general requirements of every good test: validity, reliability, and practicality. The test developer should also ensure that testing procedures give all examinees, including those with special needs, a fair and equal opportunity to be tested on their skills and knowledge.

Test Validity

The validity of a test means the extent to which a test measures what it is intended to measure. The purpose or intent of a test, in turn, is always to relate to some criterion in the real world. When the test is first conceived, this intent is reflected in the careful selection of content. When the test is tried out in preliminary form, those items that seem to measure the criterion best are determined, and only those are included in the final form. An approach for validating competency tests is described in Validating Competency Tests and Using Test Results, Module 20 in the VECS series.

Test Reliability

Test reliability refers to the consistency of the test. A reliable test would yield close to the same score for the same individual time after time. We can't expect any test to produce exactly the same results each time for the same subject even if the subject learned nothing between the first and second testing. Guessing and other factors will influence scores on repeated administrations, but a well-constructed test should yield essentially the same score each time. Thus, the relative standing of any group of examinees given a reliable test will vary only a small amount between different administrations.

Test Practicality

To be practical, a test should be feasible to construct, capable of administration without confusion, and easy to score with precision (see Adkins, 1974). This advice is absolutely vital when applied specifically to vocational competency testing.

Construction. The two most basic necessities for constructing a test are time and expertise. An adequate and realistic amount of time should be budgeted for a test development project. The development strategy and scope of the test must be within the organizational capability and personnel expertise of the test development group.

Administration. The administration of a test involves the examiner and examinees. In order to properly administer a test, the examiner requires clear, simple, and complete directions regarding every required task. The examinees must also know what is required of them. The internal simplicity and organization of a test depends on the complexity of the occupation and the level of skill to be assessed by the test. A reasonable rule of thumb is to maintain the same level of technical complexity in the test as in the occupation being tested.

Scoring. Scoring procedures that are straightforward and can translate easily into usable summaries of student performance are clearly the most useful for vocational training.

These three areas of practicality may overlap, but by considering them independently we are able to get a clearer picture of the "practical" considerations involved in test development.

Testing Individuals With Special Needs

The purpose of a vocational competency test is to assess job-related skills and knowledge. To meet this objective, it is necessary to plan the testing procedures to ensure that all examinees have a fair and equal opportunity to be tested on their skills and knowledge. Judgments about the competencies of persons with special needs should be based on their knowledge of the job and their capacity to accomplish important job-related tasks.

A helpful guide suggested by the National Research Council for use in modifying tests is the Guide for Administering Examinations to Handicapped Individuals for Employment Purposes

(Heaton, Nelson, & Nester, 1980). After a series of modifications to the testing procedures have been proposed, it is useful to have the modifications reviewed by a group of experts who are themselves handicapped. As the Council stated in its recommendations,

> No one knows as well as a knowledgeable and sensitive blind person, for example, what difficulties other blind people are likely to encounter on a particular test (p. 135).

Modifications to the testing procedure will also be necessary for persons with limited ability in English, unless knowledge of English is a requirement for performing the job satisfactorily.

Table 1 lists some adaptation suggestions for the test developers and the examiners; in addition, other reasonable local adaptations should be considered.

Judgments about the competencies of persons with special needs should be based on their knowledge and abilities, not their physical or linguistic limitations. Assessing abilities, therefore, should be determined from the outcomes rather than from the particular method used to achieve these outcomes.

TABLE 1

SUGGESTED ADAPTATION PROCEDURES FOR SPECIAL GROUPS

Special Condition	Testing Problems	Adaptations
Hearing impairment	Can't hear oral instructions	Provide printed equivalent; use audio amplification
Orthopedic upper limb disability	Can't complete response sheet blanks; difficulty in reaching or handling standard tools or equipment	Have assistant complete blanks; give test orally; permit the use of jigs and guides
Blindness/low vision	Can't read printed test materials; can't use charts and illustrations; can't see dials or markings	Read aloud and repeat; adjust testing time; prepare raised-line tactile drawings with oral descriptions; use adapted equipment (e.g., use measuring tools that have raised tactile markings); use actual objects
Limited-English proficiency	Can't understand or read directions in English	Use translations unless knowledge of English is a job requirement

Individual Study Activities

1. Select one of the three general requirements of every good test--validity, reliability, practicality--and write a one-page paper stating its importance in developing vocational competency tests. Refer to the Resources or Recommended References in this module, or to one of your own resources for information on test validity, reliability, and practicality.

2. Locate an individual with special needs (disabled or limited-English proficient) and interview that individual regarding the experiences he or she has had in being tested for vocational occupations. Did the tests provide a fair assessment of that individual's skills and knowledge? What adaptations, if any, were made in the testing procedures? Summarize your findings and share the information with the class.

Discussion Questions

1. What kinds of problems is a visually impaired person likely to face in being tested for the occupation of computer operator? What kinds of testing adaptations might be made for this individual? After the discussion, consider what you learned about your own and other classmate's attitudes toward testing individuals with special needs.

2. Discuss the implications of the following statement: "Validity is the first requisite of any test. No matter how satisfactory in all other respects, an instrument that does not provide to the decision maker accurate information of the type needed is worthless" (Erickson & Wentling, 1976, p. 22).

Group Activity

1. Divide the class into small groups, with each group representing a different category of special needs. (For example, one group may select to be hearing impaired individuals. Another group may select to be limited-English speaking.) Roleplay the problems your group would face in being tested for a vocational occupation. Create your roleplays around actual problems you know about from your own experience or have heard about from other individuals.

GOAL 2: Summarize important considerations in designing initial test specifications.

What Do You Include in Initial Test Specifications?

The initial test specifications serve as the general blueprint. In designing this blueprint the test developer must consider the purpose of the test and the limitations of the "environment" in which the test will be used. Some limits that should be considered are:

- The amount of time that can reasonably be expected for administering a test to examinees
- The availability of equipment and materials for performance testing
- The grade or mastery level of the typical examinee

The topics covered in an outline of test specifications are:

- The types of measures to be used (paper-and-pencil, performance, or both) and their formats
- The total number of items in the finished form
- Total testing time
- The skill level to be assessed by the test
- General reading level of the instructions and questions

As the test is developed, various changes are likely to be made, but the specifications should provide guidance and coherence to a test development project.

Types of Measures to be Used and Their Formats

The two types of measures considered most useful for a vocational competency test are paper-and-pencil tests and performance tests. To fully assess all skills taught in a vocational education program, you will likely need to develop a test package made up of both paper-and-pencil and performance

tests. Both tests have strong and weak points which should be carefully considered when deciding if or when to use one measure or the other.

Paper-and-pencil tests: Their strengths and weaknesses. By far the most common form of testing is paper-and-pencil. Its popularity as a test format is largely based on its flexibility, low cost, and ease of administration and scoring. A paper-and-pencil test typically requires no special equipment or specially trained staff. With a paper-and-pencil test it is possible to test a sizable group of individuals at one time, making an effective use of classroom time.

Paper-and-pencil tests do have drawbacks which can be significant in vocational education. With a paper-and-pencil test we can assess whether a student knows how to do a task, but we have little information about whether a student actually can do a task. A written test can also distort or bias our assessments. For example, in assessing a person's knowledge of small engine repair with a paper-and-pencil test, we are at the same time assessing the person's knowledge of English, reading skill, and skill in following written instructions. Overall, however, paper-and-pencil tests are still the most generally useful and practical means of assessing job knowledge.

Performance tests: Their strengths and weaknesses. A performance test in vocational education involves the examinee carrying out the actions that are expected to be performed "on the job." A performance test can be structured in the form of a simulation, copying a work situation with something less than perfect fidelity, or as a work sample which usually involves an actual "slice" of the job.

Motor skills and interpersonal skills, such as dealing with customers, are competencies which paper-and-pencil tests cannot usually assess. Since most vocational educators are concerned with "hands on" performance, a performance-based test is appropriate for assessing most of the skills taught in a vocational education program.

If, for example, an instructor of an auto mechanics course wants to determine whether the students can install a piston correctly, the most direct way of finding out would be to have each student install a piston. A performance test used in a standardized vocational competency test is simply the "teacher approach" with standardized procedures for conducting and assessing student performance. An advantage of this approach is that there is no intermediate task between what students are trained to perform and how they are assessed.

Though this approach has significant advantages for assessment in vocational education, there are important drawbacks:

1. Time--Performance tests are more time consuming. Often they must be administered on a one-to-one basis (one examiner to one examinee), with time required between each examinee for setting-up the test.

2. Cost--Special equipment and materials are needed and often consumed during testing.

Because of the time and cost constraints, performance tests may be limited in the range of tasks that can be assessed. Consequently, for maximum efficiency, they should be used only for those competencies that cannot be assessed adequately by paper-and-pencil tests.

Common formats of paper and pencil tests. The four most common paper-and-pencil test formats are: true-false, matching, completion, and multiple-choice.

1. True-false--True-false test items have an advantage in ease of construction but should be limited to factual material. Examinees can usually answer a large number of these items in a relatively short period of time, allowing a broader sampling of knowledge than is possible with other formats within the same time period. Even when testing factual information, the true-false format has serious disadvantages. Random or "blind" guessing will allow a person, on an average, to respond correctly to 50 percent of the items. This high guess-factor could allow someone who knows nothing about the subject to get half the items correct, thus negating any advantage gained by asking a large number of questions.

2. Matching items--Matching is a format in which questions are arranged in one column and alternatives or answers are in a second column. Examinees are asked to select the correct response in the second column that corresponds to the question, word, or statement in the first column. This format can be made to have a high difficulty level and reduce the proportion of correct answers through guessing. The questions must be closely related so that, for any question, the incorrect choices will serve as reasonable distractors. The listing of items should be kept relatively brief (10 to 15 alternatives) so that finding the

correct response does not become tedious. When making up the two lists for matching, it is recommended that one column contain several more items than the other. This lessens the chance of selecting answers based simply on the process of elimination.

3. Completion items--Completion items (sometimes referred to as "fill in the blanks") are deceptively easy to construct and have no significant guess factor. Scoring, however, is difficult since no specific responses or options are given, and responses cannot be machine-scored. A subject matter expert may be needed to determine the correctness of every unanticipated response. Such a scoring procedure is slow, expensive, and borders on the subjective. Under conditions where there are no guarantees of the expertise of the scorer or where a large number of tests are to be administered, this format is not practical.

4. Multiple-choice items--A multiple-choice format is usually the most desirable format for knowledge items in a vocational competency test. In general, a multiple-choice item consists of the stem, which may be either a complete or an incomplete statement, and several responses that answer the question directly or complete the statement. Of the responses given, only one is correct, and the rest serve as distractors (see Table 2).

TABLE 2

COMPONENTS OF A MULTIPLE-CHOICE ITEM

An engine that fires each time the piston goes up is a	STEM
A. 2-cycle engine.	CORRECT RESPONSE
B. 4-cycle engine.	DISTRACTORS
C. rotary engine.	DISTRACTORS
D. supercharged engine.	DISTRACTORS
E. turbocharged engine.	DISTRACTORS

The number of responses usually varies between three and five. If three responses are used, the possi-

bility of guessing correctly is about 33 percent; for four items, 25 percent; for five, 20 percent; and so forth. Five responses are recommended as giving an appropriate balance between item length and the effects of guessing.

Constructing distractors (the incorrect responses) is often the most difficult part in writing multiple-choice test items. Each distractor must:

- agree grammatically with the stem,
- be rational and believable,
- be related to the topic area,
- NOT contain any hints to the correct answer,
- seem plausible and attractive to the uninformed or poorly prepared examinee, and
- BE ABSOLUTELY WRONG.

The use of responses such as "all of the above," "A and B but not C," etc., should be avoided in most cases. These types of responses can be useful under some circumstances, but they can easily become crutches for item writers.

Although the multiple-choice test item format is the most widely used one in large-scale testing efforts, other possible formats should not be discarded. It is also common to mix test formats, if for no other reason than to break the monotony involved in testing. The test designer should use the most appropriate format for each situation and weigh the benefits and drawbacks inherent in each.

Types of performance evaluation. Two types of evaluation procedures usually used for performance testing are product evaluation and process evaluation.

1. Product evaluation--Evaluating a product is usually the simpler of the two because the examiner evaluates the results of actions and not the actions themselves. Products can be measured or checked against specific standards anytime after the examinee has completed the test. Such things as a completed electrical circuit or a carburetor rebuilt to factory specifications are examples of products that can be evaluated.

2. Process evaluation--A process evaluation means that actions and behaviors are assessed while the activity is in progress. For example, the evaluation of meal service by a waiter, or lifting a patient properly from a bed by a nurse's aide, are evaluations of the process performed. An actual product may or may not result. The actions are viewed as the most important component of the job task.

The type of performance evaluation used will depend on what is being assessed. Often, a performance test is partly process and partly product based.

Individual Study Activities

1. Select an occupation from your area of occupational specialty and develop an outline of test specifications for that occupation. Consider whether you will use standardized tests or develop your own. Consider the types of measures you will use (paper-and-pencil, performance, or both) and their formats; the total number of test items; total testing time; the skill level to be assessed by the test; and the general reading level of the instructions and questions.

2. Using the Resources or Recommended References in this module, or one of your own resources, develop two charts--one listing the strengths and weaknesses of paper-and-pencil tests and the other listing the strengths and weaknesses of performance tests. Then compare the two charts and determine which type of test has the greatest overall strengths to meet the needs in your particular setting.

Discussion Questions

1. "Measurement techniques for assessing occupational students' achievement within the cognitive domain have been used much more extensively than those used to assess achievement in the affective, psychomotor, and perceptual domains" (Erickson & Wentling, 1976, p. 86). Using this statement as a basis of discussion, provide examples from your own experience that support this statement.

2. "The ultimate in performance measurement for occupational education is the assessment of a student's ability to perform important job-related tasks in an actual job setting" (Erickson & Wentling, 1976, p. 126). This type of measurement, however, has not been used extensively by vocational educators. Discuss why this is so and suggest ways that performance tests might be used more.

Group Activity

1. Divide the class into four groups, each group representing one of the four types of paper-and-pencil test formats: true-false, matching, completion, and multiple-choice. Each group will defend the use of its particular format in a vocational competency test.

GOAL 3: Discuss the critical tasks in developing paper-and-pencil tests.

How Do You Develop Paper-and-Pencil Tests?

Developing paper-and-pencil tests is a systematic process consisting of a number of specific tasks. A discussion of these tasks follows.

Create An Item Budget

The "budgeting of items" is a procedure for determining the number of paper-and-pencil test items to develop within each major test area. Your decisions about an item budget must be based on your best estimate; no hard and fast rules exist.

The first step is to establish the content of the item budget, that is, what will be the areas of an occupation to be covered in the test. The content is derived from the task inventory findings listing the skills and abilities that employers and employees consider important. (See Module 18: Determining Requirements for Vocational Competency Measures for a discussion of developing a task inventory.)

The next step is to estimate the number of items for the final test. For example, if between 50 and 60 test items are wanted for the final version, at least double that number of items should be prepared initially. The test developer must then assign a percentage of the total number of test items being developed to each major category. Examples of major categories derived from an auto mechanics task inventory could be "safety," "trouble-shooting," and "tune-up procedures."

If you decide to prepare 100 items initially, and the test will have five major categories, you could assign the same number to each category. In this case, you simply have:

$$100 \div 5 = 20$$

Often, some categories are more important to the job than others; you may wish to assign a larger proportion of items to

those categories. The distribution, then, could resemble the following:

Category	No. of Items Per Category
A	30
B	10
C	20
D	25
E	15
TOTAL	100

Once the numbers of test items to be developed are assigned to the major categories, the number of items to be developed for each individual area or task within each major area in the task inventory must be determined. If, for example, a major category is allotted 30 questions, such as category "A" above, and if the category has five specific tasks, then the 30 questions must be assigned to these five tasks.

Often, tasks to be used as the basis for item development are rated by those surveyed as to their importance. If, for example, tasks were rated either "moderately important" or "very important," such that:

Assigned Weight

1 = moderately important
2 = very important

then the first category could be assigned "1" and the second, "2." We could assign test items per task using the following simple procedure. We sum the total weight of all the tasks:

Task	Importance		Assigned Weight
1	moderate	=	1
2	very	=	2
3	very	=	2
4	moderate	=	1
5	moderate	=	1
			7

We then divide the allotted number of questions, 30, by the total weight, 7:

$$30 \div 7 = 4.3$$

The result, 4.3, is then multiplied by the weight of the task, which in this case is either 1 or 2.

Task	Assigned Weight		Weight Multiplier		No. of Questions Per Task
1	1	x	4.3	=	4.3
2	2	x	4.3	=	8.6
3	2	x	4.3	=	8.6
4	1	x	4.3	=	4.3
5	1	x	4.3	=	4.3

Since these results are not whole numbers, it is necessary to round to the nearest whole number to get the appropriate number of questions to develop per task, so that the final result is:

Task	No. of Questions Per Task
1	4
2	9
3	9
4	4
5	4
TOTAL	30

Tasks that are different aspects of the same general skill are best grouped so as not to overtest in one area at the expense of other tasks. In retail sales, for example, "prepare a sales receipt" and "prepare a refund form" can easily be combined since the skills involved are essentially the same.

Assign Test Items

Based on the test item budget, assign each item writer a specific number of test items to write in specific categories. Supply each writer with detailed instructions regarding the nature of the task, specific requirements, and the format of the questions to be developed. Table 3 is a listing of areas that might be covered in instructions to item writers. These instructions should clearly reflect the design goals of the test developer.

Of course, items should be assigned to writers who are knowledgeable about the specific tasks. But it is a good idea to ask a second writer to submit items on the same tasks to vary the emphasis and point of view.

TABLE 3

SUGGESTED INSTRUCTIONS TO TEST ITEM WRITERS

- Establish a time limit for item preparation.
- All items should be sent to the test development center (give address).
- After typing and editing at the development center, all items will be returned to the writer for review to make certain that editorial changes have not changed the technical content.
- For each item, indicate the correct answer and the topic to which the item is related.
- Staff members will meet with item writers as appropriate to discuss problems or concerns, and to review reasons for editorial changes.
- Five-choice, multiple-choice items are preferred. Items with fewer than five choices will be accepted if there is good reason for the smaller number of choices.
- The stem of each item should be a complete question or an incomplete statement. Stems with a blank in the middle should be avoided.
- All options must relate to the stem logically and grammatically.
- Discourage the use of options such as "All of the above," "None of the above," and "A and B above."[1]
- Stems and options should be as short as possible while still being complete.
- Never repeat a word in the options if it can be included in the stem.
- Options should be arranged in some logical order such as:

  Numerical
  Alphabetical (for one- or two-word options)
  Length (for multi-word options)

- Avoid the use of words such as "always" and "never."
- Make certain that level of reading difficulty is appropriate to audience.

[1] Though options of this type may at times prove useful, item writers should be initially discouraged from their use because they are apt to be misused. If, for example, item writers are asked to prepare five-option multiple-choice items, it becomes extremely tempting to have a fifth option as "All of the above" or "None of the above" because it is an easy distractor to write.

Review and Modify Items

The test development staff should review each completed item for clarity, completeness, correct grammar, etc. Items that are unclear, unimportant, or have content problems requiring technical expertise should be discussed with the item writer. Occasionally, when different individuals prepare items for the same occupational task, duplicate or near duplicate items will be written. You can keep the best item and place the other in reserve, or possibly even combine the two to produce a better test item.

Then compare the selected items to the item budget. You may find that you have more stems than you need in one area and too few in another. In areas of shortage, the test developer should work closely with the item writers to stimulate ideas for questions. A "brainstorming" session can be very productive. Once there is a close match with the budget, the items should be prepared in final form. The language and structure of the questions and their options should be logically and grammatically consistent. All the items should then be looked at as a unit so that extraneous cues can be removed. A common error is that one test item will cue the correct response in another item.

The items should then be reviewed by subject matter experts not involved in the actual writing. You might have vocational instructors or persons in industry not previously involved act as reviewers. Be sure to remove items that are factually in error or items in which a disagreement exists about the correct response. A good strategy is to have the experts "take the test." Although even experts can make errors, it should be possible to catch most of the distractors that may be correct responses to the question.

Table 4 lists questions that should be considered in reviewing each item.

TABLE 4

AREAS OF CONCERN WHEN REVIEWING PAPER-AND-PENCIL TEST ITEMS

- Is the content of the item logical?
- Is sufficient information provided?
- Is the use and spelling of technical terms correct?
- Is there one and only one best answer?
- Is the correct answer keyed?
- How difficult is the item apt to be for a typical student?
- Can the item be improved? If so, how?

Balance the Key

Balancing the key is one of the final stages in the preparation of multiple-choice test items. It ensures that each possible option has approximately an equal proportion of correct answers assigned. It helps to eliminate any kind of bias that test designers or item writers may have in the placement of correct answers. Item writers have a tendency to plan the correct response as one of the middle options.

The balancing procedure to use is very simple:

1. List separately the correct response for all five-choice items, four-choice items, and three-choice items.
2. Sum the number of five-choice, four-choice, and three-choice items separately.
3. Divide each sum by the number of choices for that category.
4. The result will tell how many A, B, C, etc., correct responses should be in each category.

5. Compare the results against the actual number of items in each category.

6. Rearrange the options, whenever feasible, to match the balanced key.

The following is an example of this balancing procedure, using five-choice items. The same procedure is used for four- and three-choice items.

Item Choice	Present Key	Adjustments Needed	Revised Key
A	6	+2	8
B	12	-3	9
C	18	-9	9
D	1	+7	8
E	5	+3	8
	TOTAL 42		TOTAL 42

Number of items ÷ number of choices per item = 42 5 = 8.4

We need about 8 questions having each item choice. In this case since the total is 42, two item choices must have one extra each.

When balancing the key, you cannot simply move options around at will. The option item ordering guidelines described in Table 3 should be maintained.

Prepare Test Administration Instructions

The best test can be a poor assessment tool if the instructions for its administration are not clear and specific. One method of providing these instructions is to prepare an examiner's manual. This manual should contain all the information required for administering the test. The manual is also useful in supplying background information to aid the examiner in understanding the overall structure and purpose of the test. Table 5 lists content areas that might be covered in an examiner's manual.

TABLE 5

SUGGESTED CONTENT AREAS OF AN EXAMINER'S TEST MANUAL

- Purpose of the test
- Expected background of examiner
- Ways test results may be used
- Overview of the test
- Administration of the test
  - test format
  - time required
  - answer sheets
  - instructions for administering the test
  - suggestions for testing individuals with special needs

Review and Revise (The Second Time)

When the test is completed, the entire package should be reviewed by the members of the test development consultant panel. To ensure useful reviews, detailed instructions should be supplied to each reviewer. Table 4 can be used as a model for the content of instructions to test reviewers.

Test revision based on the comments of review panel members should be carefully performed to avoid introducing errors into the test. Any changes of factual content that seem questionable should be submitted to the item writers. Carefully review any area where there are differences of opinion. If it is not possible to agree on the correct answer, give serious consideration to deleting the item.

Conduct a Pilot Test

Once the test has been revised, it is ready for pilot testing. Pilot testing is a procedure used primarily to test the structure of the test, e.g., clarity of instructions, time estimates, etc., rather than the content. The procedure for conducting a pilot test is to have two or three vocational training programs each administer the test to two or three students. The tests should each be at a different school with different examiners and students. A packet of the complete test instructions should be sent to each examiner, one or two weeks in advance.

The examiner should be informed that the test designer will be at the test site as an observer and to debrief the examiner and examinees. Make it very clear that the test itself is being tested and that any problems with the test are not a reflection on the examiner or student but rather on the test. Ask the examiner and the examinees to be completely frank and not worry about hurting the feelings of the test designer.

During testing, the test designer will act as an observer and should not supply assistance or answer questions. It is appropriate, however, to stop and ask questions for clarification. Careful notes should be kept, noting the strong and weak points of test administration.

After completion of the test, the observer should review the test with the examiner and students. An item-by-item review is most desirable. On completion of the pilot tests, corrections should be made to the test based on the findings. The findings of pilot testing will help improve instructions and "packaging" of the test. Once these problems are corrected, you are ready for field testing.

Conduct a Field Test

Field testing is used to determine the quality of the test items. The field test involves a large number of students in a number of schools. Field testing requires considerable preplanning in that agreements must be obtained from the schools and instructors. There are no hard and fast rules regarding the size of the sample. Approximately 100 students spread over five schools is a reasonable target.

Analyze Field Test Findings

Field testing will provide the test developer with a large amount of data to analyze. If possible, the developer should work closely with someone skilled in statistical analysis. To do a good job of analysis, access to a computer is necessary. If you have access to a computer, most likely you will find one of the numerous statistical software packages in place. Most are simple to use and can accommodate the level of statistical analysis required for item analysis purposes.

Revise the Test

In order to decide which items to keep and which to discard, it is necessary to get statistical data on each item to answer the following questions.

1. How well does examinee performance on each item correlate with overall examinee performance on all items in a particular subject matter area? (How well does an item discriminate between those who perform well and those who perform poorly overall?)

2. What was the difficulty level; that is, what proportion of examinees answered an item correctly?

Your prime concern should be to select items that performed well in the field test and to end up with a test that has about the same percentage of items in each section as specified in the item budget. Your criteria for item selection should be threefold:

1. Content--The items should be representative of the performance area.

2. Difficulty--What was the proportion of examinees that missed any item? In competency tests, items should be included over a range of difficulty to reflect a range of competency.

3. Discrimination--Do those who do well on the test generally answer the item correctly and those who do poorly overall generally answer the item incorrectly?

In analyzing the paper-and-pencil items, consider these points. Distractors that were chosen by few examinees might need to be modified or deleted. Items answered correctly by all examinees should be eliminated or changed. Items which are seldom answered "correctly" should be checked for keying.

If the key is correct, they should be carefully reviewed by subject matter experts to determine whether there is a possibility of misinterpretation.

After you have selected test items that are acceptable on the basis of content, difficulty, and discrimination, make sure you have the same item distribution as you had on the field test. If some areas do not have enough items, look again at those items not selected.

Generally, poor items should be removed rather than rewritten, but if the change is small the test developer may want to make it in order to have a useful item. If a major change is required, then the rewritten item should be considered a new item that has not been field tested.

On completion of the final revision, you should have a useful paper-and-pencil test meeting the design requirements established at the start of the project.

Individual Study Activities

1. Obtain a teacher-made occupational competency test from your area of occupational specialty. Randomly select 10 multiple-choice, paper-and-pencil test items and review them to answer the following questions: Is the content of the item logical? Is sufficient information provided? Is the use and spelling of technical terms correct? Is there one and only one best answer? How difficult is the item for a typical student? Can the item be improved? If so, how? Revise the test items based on your review.

2. Obtain a task inventory for an occupation from your area of specialty. Select two of the very important tasks and write five multiple-choice test items for each task. Follow the suggested guidelines in Table 3 of this module. Then exchange your items with a partner and review one another's items based on the review questions above. Return the items to the writer and revise them according to your partner's suggestions. You must determine, of course, whether the suggestions are appropriate.

Discussion Questions

1. This module discusses a number of specific tasks for developing paper-and-pencil tests. How practical are these tasks for your particular setting? If the members of the class were part of a district-wide test development team, which of the test development tasks would you expect to be able to accomplish? Write these tasks on the chalkboard and reach a consensus regarding the tasks you will do.

2. Why do you think it is necessary to develop the examiner's test manual as part of a total test package? What topics would you include in such a manual and why?

Group Activity

1. Organize class members into a test development team. Select a leader and assign responsibilities among members. Your goal is to produce a paper-and-pencil vocational competency test for an occupational area of need in your district or state. If all the skills necessary for test development are not represented on your team, indicate what skills are lacking and where you would obtain that expertise.

GOAL 4: Discuss the critical tasks in developing performance tests.

How Do You Develop Performance Tests?

Many of the tasks required for performance test development can be carried out in conjunction with those required for the development of paper-and-pencil tests.

The following five components should be covered in each performance test:

1. Purpose--Statement giving overview of the task, the specific subtasks involved, and the uses that can be made of the findings. This statement is for the use of the examiner.

2. Instructions to examiner--Written descriptions/instructions of exactly what an examiner is expected to do step-by-step during all aspects of testing. The instructions should be detailed and specific.

3. List of required equipment--Detailed description of the exact layout of the test site and all required materials.

4. Instructions to examinee--Written instructions either read by or to the examinee. They must be brief yet complete, giving no extraneous information that could distract the examinee from the task at hand.

5. Rating form--Document that lists checkpoints for assessing the job competency of the examinee on the specific task being performed. Each checkpoint should have a rating checkoff on which the examiner can quickly record the correctness or acceptability of the examinee's performance.

Select the Tasks to be Tested

Performance tests are more difficult to construct and more expensive to administer than paper-and-pencil tests. Therefore, performance tests should be developed to cover only important areas that cannot be tested adequately using a paper-and-pencil test. It is important to discuss each proposed

area with subject matter experts. Some questions to ask in selecting tasks for a performance test are:

1. Can the skills required to perform the task only be assessed adequately with a performance test and not with a paper-and-pencil test?

2. Was the area rated as important by employers and employees in the task inventory survey?

3. Are the equipment or required materials available or easily obtainable at a training site?

4. Will the cost in consumable items per examinee be reasonable?

Structure the Tasks

As you can see from the list of components of a performance test, the final package is much more complex than a paper-and-pencil test package. It is the designer's job to make all the parts fit together so that those using the test feel it is simple to set up, administer, and score.

First, the task should be divided into all its observable behaviors and products. For example, if an examinee is asked to replace the head gasket of an automobile engine, the task should be broken down and each step recorded. This includes even those actions that seem trivial. In the case mentioned, we would start with:

1. Examinee opens hood of car.

2. Determines model of engine.

3. Checks to see whether proper gasket is available.

The listing would be finished when the examinee closes the hood of the car. The list should be ordered as closely as possible to the way the task is typically performed. The test developers should then look for possible products resulting from the process. In our example, such points as "heads torqued to proper specification," "gasket straight and not leaking" are potential products that can be measured.

From the list, those points that are (a) important to the job, (b) likely to be performed wrong, and (c) relatively independent of other actions required for performance of the task should be extracted and used as assessment checkpoints. The

checkpoints should then be written as descriptions of the action, and in enough detail to ensure that an examiner can make a judgment about correctness of a behavior/action. Too much detail can make it difficult for an examiner to rate an action in an ongoing work situation.

Table 6 is an example of checkpoints at the proper level of detail for one important task in overhauling a diesel engine.

Prepare Administration Instructions

To ensure that a performance test is administered the same way to all examinees, detailed administration instructions must be prepared. An examiner's manual should cover all the components of a performance test. It should also include suggestions for testing individuals with special needs. Though the instructions to the examiner should be detailed enough to prevent variation between tests, they should not be unnecessarily wordy.

Develop Rating Sheets

Rating sheets on which the examiner will rate examinee performance should be easy to follow and at the same time permit a thorough evaluation of the examinee's performance. The rating form should allow an examiner to observe and rate the examinee on each checkpoint without missing any of the examinee's ongoing performance on the test. Table 7 is an example of a rating sheet derived from the checklist shown in Table 6. Note that product checkpoints 12-16 have been added to the initial process checkpoints.

Review by Consultants

As with paper-and-pencil tests, performance tests should be reviewed by subject matter experts in industry and by vocational educators. The points that reviewers should keep in mind are listed in Table 8. If a paper-and-pencil test is developed, it would be most efficient to have both the paper-and-pencil and performance tests evaluated by the same individuals at the same time. The reviewers can then look at the complete test package.

TABLE 6

CHECKPOINTS USED IN THE DEVELOPMENT OF A PERFORMANCE TEST FOR DIESEL MECHANIC

Performance Test: Measure Gap and Side Clearance and Install Piston Rings	
Task	Performance Criteria Checkpoints
Measure ring end gap clearance.	1. Measures each ring individually. 2. Uses piston to push each ring into sleeve. 3. Makes sure rings are straight in sleeve. 4. Uses feeler gauge to measure gap.
Measure ring to groove side clearance.	5. Holds each ring against proper groove on piston or mounts rings in proper grooves. 6. Uses feeler gauge to measure side clearance of each ring.
Install rings on piston.	7. Uses expanders to install each ring on piston. 8. Installs each ring in proper groove.
Prepare for installing piston in sleeve.	9. Staggers ring gaps. 10. Insures that no ring gap is in line with wrist pin hole. 11. Cleans and puts tools away.

TABLE 7

A SAMPLE RATING SHEET FOR DIESEL MECHANIC

Performance Test: Measure Gap and Side Clearance and Install Piston Rings

Performance Test Record Sheet

Examinee_____ Examiner_____ Date_____ (No. Day Yr.)

School/Employer _____

Examiner must enter the correct measurements and the specifications from the manual for items 12-21.
Evaluate the examinee's performance on the following tasks by checking either the "Yes" or "No" column.

Task	Performance Criteria		Yes	No
Measure ring end gap clearance	1. Measures each ring individually	1.		
	2. Uses piston to push each ring into sleeve	2.		
	3. Makes sure rings are straight in sleeve	3.		
	4. Uses feeler gauge to measure gap	4.		
Measure ring to groove side clearance	5. Holds each ring against proper groove on piston or mounts rings in proper grooves	5.		
	6. Uses feeler gauge to measure side clearance of each ring	6.		
Install rings on piston	7. Uses expanders to install each ring on piston	7.		
	8. Installs each ring in proper groove	8.		
Prepare for installing piston in sleeve	9. Staggers ring gaps	9.		
	10. Insures that no ring gap is in line with wrist pin hole	10.		
	11. Cleans and puts tools away	11.		
Top ring: gap	12. Manual specification recorded by examinee (_____) matches manual specification (_____)	12.		
	13. Gap measured by examinee (_____) matches gap measured by examiner (_____)	13.		
Top ring: side clearance	14. Examinee correctly checks "NO" (ring does not need replacement)	14.		
	15. Manual specification recorded by examinee (_____) matches manual specification (_____)	15.		
	16. Side clearance measured by examinee (_____) matches side clearance measured by examiner (_____)	16.		
Second ring: gap	17. Manual specification recorded by examinee (_____) matches manual specification (_____)	17.		
	18. Gap measured by examinee (_____) matches gap measured by examiner (_____)	18.		
	19. Examinee correctly checks "NO" (ring does not need replacement)	19.		
Second ring: side clearance	20. Manual specification recorded by examinee (_____) matches manual specification (_____)	20.		
	21. Side clearance measured by examinee (_____) matches side clearance measured by examiner (_____)	21.		

TABLE 8

CHECKLIST FOR REVIEWING PERFORMANCE TEST ITEMS

- Is the situation realistic?
- Are all required supplies and materials listed? If not, what is missing?
- Are the instructions adequate?
- Are all the relevant topics for evaluation included? If not, what topics are missing?
- Are the topics for evaluation listed in the order in which they would be carried out by an examinee taking the test?
- Are the materials provided sufficiently clear and complete? If not, what should be added to make them more satisfactory?
- Are the approximate time limits indicated in the test satisfactory?
- How can the problem be improved?

Test the Performance Test

The procedures for pilot testing and field testing a performance test are basically the same as those used for a paper-and-pencil test.

The individual performance tasks should correlate positively--but not necessarily highly--with the job knowledge test. If one or more correlate very highly, that raises a question of the cost effectiveness of these performance measures. Another concern is if a performance test has a negative correlation with the paper-and-pencil test. The correlations among performance measures should also be reviewed. Ideally, they will have low positive correlations.

Within each performance test, checkpoints that are performed correctly by all or nearly all of the examinees should be carefully reviewed. The importance of the item to the occupation and the task should be considered. Checkpoints that are missed by all or nearly all examinees should be restudied to determine what is causing the problem.

As part of final revision, the test developer should spend time working on the appearance and layout of the test. Every effort should be made to reduce the amount of paper and paperwork. The final version of the test should be clear, well-organized, and easily administered. At this point, the developer should have a complete test package that will give objective information to directors of vocational education, instructors, and students about the capability, strengths, and weaknesses of students and programs.

Individual Study Activities

1. Obtain a teacher-made performance test in an area of your occupational specialty. Review the test to answer the following questions: Is the situation realistic? Are all required supplies and materials listed? If not, what is missing? Are the instructions adequate? Are all the relevant topics for evaluation included? If not, what topics are missing? Are the topics for evaluation listed in the order in which they would be carried out by an examinee taking the test? Are the mateials provided sufficiently clear and complete? If not, what should be added to make them more satisfactory? Are the approximate time limits indicated in the test satisfactory? How can the test be improved?

2. Obtain a task inventory for an occupation from your area of specialty. Determine which tasks would be appropriately assessed by a performance test. Some questions to ask yourself are: Can the skills required to perform the task only be assessed adequately with a performance test and not with a paper-and-pencil test? Are the equipment and required materials available or easily obtainable at a training site? Will the cost in consumable items per examinee be reasonable? Provide reasons for your selection of tasks for performance testing.

Discussion Questions

1. "In assessing student performance in occupational education programs, either or both the process or product of the task should be measured. Both product and process assessments have their advantages--a decision must be made regarding which should be used" (Erickson & Wentling, 1976, p. 128). From your own experiences, can you think of situations in which it would be advantageous to look at both process and product? Discuss these situations.

2. "The common conception that paper-and-pencil tests of performance can only measure cognitive functioning is not entirely true. Many paper-and-pencil tests can provide direct assessment of job performances" (Erickson & Wentling, 1976, p. 155). From your own experiences, can you think of examples of a written test that in fact serves as a performance test? Discuss these examples.

Group Activity

1. Organize class members into a test development team. The goal of the team is to produce a performance test for an occupational area of need in your district or state. As team members, determine which specific tasks you will reasonably be able to carry out in your particular setting to achieve this goal. Prepare an action plan for developing the performance test, indicating tasks and individuals responsible for carrying out those tasks.

Summary

The procedures described in this module, in combination with the other three modules of this series, will help you develop a useful and valid vocational competency test. This module has focused on techniques and procedures essential to a good test development effort.

Another component of equal importance, but beyond the scope of this module, is teamwork. Developing a vocational competency test is not a one-person job. It requires the input of many people. Without this input, the developer would probably have only a "classroom test made large" and not a test that realistically measures occupational competency as determined by business and industry.

A test designed following these procedures will serve as more than simply an assessment of student performance. It will permit programs to be evaluated on the basis of those skills desired by business and industry.

We feel vocational training programs should not focus on only those areas covered in a test. For the very practical considerations of time and budget, a test can only be a sample of the topics included in a good training program. Nevertheless, the results of a good test assessing important competencies will definitely contribute to improving vocational education programs.

APPENDICES

Self-Check

GOAL 1

1. What is the importance of test validity, test reliability, and test practicality in vocational competency test development?

2. Why is it important to develop test procedures that accommodate individuals with special needs?

GOAL 2

1. What items should be included in initial test specifications?

2. For both paper-and-pencil tests and performance tests, what are two major strengths and two major weaknesses of each?

3. What are four common formats of paper-and-pencil tests?

4. What are two types of performance evaluation?

GOAL 3

1. What is the purpose of creating a test item budget for paper-and-pencil tests?

2. What are important considerations in initially reviewing and modifying paper-and-pencil test items?

3. What are the differences between pilot testing and field testing vocational competency tests?

4. After field testing, on what basis should paper-and-pencil test items be revised?

GOAL 4

1. What are the components of a performance test?

2. What is the major consideration in selecting tasks for performance test development?

3. Performance tests should be reviewed by subject matter experts in industry and vocational education. What considerations should they keep in mind when reviewing these tests?

Self-Check Responses

GOAL 1

1. Any good test should have validity, that is, it should measure what it is intended to measure.

   Any good test should have reliability, that is, it should be able to identify, with a high degree of accuracy, the relative standing of any group of examinees.

   To be useful a test must have practicality in terms of construction, administration, and scoring. It should have an adequate amount of development time budgeted, be capable of administration without confusion, and easy and straightforward to score.

2. Planning testing procedures to accommodate individuals with special needs ensures that all examinees have a fair and equal opportunity to be tested on their skills and knowledge.

GOAL 2

1. Initial test specifications should include:

   - Types of measures to be used and their formats
   - Total number of items
   - Total testing time
   - Skill level to be assessed by the test
   - General reading level of instructions and questions

2. Strengths of Paper-and Pencil Tests
   - Flexibility, low cost, ease of administration and scoring
   - Requires no special equipment or specially trained staff

   Weaknesses of Paper-and-Pencil Tests
   - Often provide little information about whether a student can do a task
   - Can distort or bias assessments if reading difficulty exceeds job demands

   Strengths of Performance Tests
   - Can assess motor skills and interpersonal skills
   - Allow students to be assessed directly on what they are trained to perform

Weaknesses of Performance Tests

- Are time-consuming
- Involve costly equipment and materials often consumed during testing

3. Common formats of paper-and-pencil tests:

   - True-false items
   - Matching items
   - Completion items
   - Multiple-choice items

4. Types of performance evaluation:

   - Product evaluation
   - Process evaluation

GOAL 3

1. The "budgeting of items" is a procedure to determine the number of paper-and-pencil test items to develop for each major test area.

2. Considerations in initially reviewing and modifying paper-and-pencil test items:

   - Is the content of the item logical?
   - Is sufficient information provided?
   - Is the use and spelling of technical terms correct?
   - Is there one and only one best answer?
   - Is the correct answer keyed?
   - How difficult is the item for a typical student?
   - Can the item be improved? If so, how?

3. Pilot testing is a procedure used primarily to test the structure of the test rather than the content. It is a small-scale test, only involving a few students in a few programs.

   Field testing is conducted to determine the quality of the test items in terms of content. It involves a large number of students in a large number of schools.

4. Bases for revising paper-and-pencil test items after field testing:

   - Content
   - Difficulty
   - Discrimination

GOAL 4

1. Components of a performance test:

   - Purpose
   - Instructions to examiner
   - List of required equipment
   - Instructions to examinee
   - Rating form

2. The major consideration in selecting tasks for performance test development is whether or not the skills required to perform the task can only be assessed adequately with a performance test and not with a paper-and-pencil test.

3. Review of performance tests by subject matter experts should include:

   - Is the situation realistic?
   - Are all required supplies and materials listed? If not, what is missing?
   - Are the instructions adequate?
   - Are the relevant topics for evaluation included? If not, what topics are missing?
   - Are the topics for evaluation listed in the order in which they would be carried out by an examinee taking the test?
   - Are the materials provided sufficiently clear and complete? What should be added to make them more appropriate?
   - Are the approximate time limits indicated in the test realistic?
   - How can the test be improved?

Recommended References

Adkins, D. C. Test construction: Development and interpretation of achievement tests (Second Edition). Columbus, Ohio: Charles E. Merrill Publishing Co., 1974.

Dahl, P. R., Appleby, J. A., & Lipe, D. Mainstreaming guidebook for vocational educators: Teaching the handicapped. Salt Lake City, Utah: Olympus Publishing Co., 1978.

Davis, F. B. Standards for educational and psychological tests. Washington, D.C.: American Pscyhological Assoc., 1974.

Erickson, R. C., & Wentling, T. L. Measuring student growth: Techniques and procedures for occupational education. Urbana, Ill.: Griffon Press, 1976.

French, J. W., & Michael, W. B. (Cochairmen). Standards for educational and psychological tests and manuals. Washington, D.C.: American Psychological Assoc., 1966.

Heaton, S. M., Nelson, A. V., & Nester, M. A. Guide for administering examinations to handicapped individuals for employment purposes. Washington, D.C.: U.S. Office of Personnel Management, Personnel Research and Development Center, 1980.

Jaeger, R. M., & Tittle, C. K. Minimum competency achievement testing: Motives, models, measures, and consequences. Berkeley, Calif.: McCutchan Publishing Corp., 1980.

Office of the Federal Register. Code of federal regulations, Labor, 29, Parts 900-1899. Washington, D.C.: U.S. Government Printing Office, 1981.

Popham, W. J. Education evaluation. Englewood Cliffs, NJ: Prentice-Hall, 1975.

Sherman, S. W., & Robinson, N. M. (Eds.). Ability testing of handicapped people: Dilemma for government, science, and the public. Washington, D.C.: National Academy Press, 1982.

Thorndike, R. L. (Ed.). Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Weisgerber, R. A., Dahl, P. R., & Appleby, J. A. Training the handicapped for productive employment. Rockville, MD: Aspen Systems Corporation, 1980.